REPORT DOCUMENTATION PAGE

AD-A198 042	**1b. RESTRICTIVE MARKINGS**
	3. DISTRIBUTION/AVAILABILITY OF REPORT *Distribution unlimited*
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report # 7	**5. MONITORING ORGANIZATION REPORT NUMBER(S)** AFOSR·TR· 88-0841

6a. NAME OF PERFORMING ORGANIZATION	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
Texas A&M University		Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code)	7b. ADDRESS (City, State and ZIP Code)
College Station, TX 77853	Same as 8

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NM	AFOSR-SF49620-85-C-0144

8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS.			
Bla 410 Bolling Air Force Base Washington, DC 20332	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
	61102F	2304	A6	

11. TITLE (Include Security Classification)
A NOTE ON EXTENDED QUASI-LIKELIHOOD

12. PERSONAL AUTHOR(S)
Davidian, M. and Carroll, R.J.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
technical Journal	FROM 8/87 TO 8/88		16

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	exponential family, heteroscedastic regression model, inference for variance parameters, pseudo-likelihood estimation, variance function estimation

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We study the method of extended quasi-likelihood estimation and inference of a variance function recently proposed by Nelder & Pregibon (1987). The estimates are inconsistent in general, and the test levels can be biased, but in many cases such as the exponential family the inconsistency and bias will not be a major concern. Extended quasi-likelihood is compared with Carroll & Ruppert's (1982) pseudo-likelihood method, which gives consistent estimates and, when slightly modified, asymptotically unbiased tests. We quantify the notion of a problem in which the amount of statistical information is large in each unit, showing in this instance that the two estimates are closely related and may be asymptotically equivalent in many important cases. However, in some cases outside the exponential family, an asymptotic bias can persist.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Major Brian Woodruff	(202) 767-5026	NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

A NOTE ON EXTENDED QUASI-LIKELIHOOD

Technical Report # 7

M. Davidian and R. J. Carroll

A NOTE ON EXTENDED QUASI-LIKELIHOOD

M. Davidian

Department of Statistics
North Carolina State University
Box 8203
Raleigh, North Carolina, U.S.A.
27695-8203

R.J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill
321 Phillips Hall 039 A
Chapel Hill, North Carolina, USA
27514

SUMMARY

We study the method of extended quasi-likelihood estimation and inference of a variance function recently proposed by Nelder & Pregibon (1987). The estimates are inconsistent in general, and the test levels can be biased, but in many cases such as the exponential family the inconsistency and bias will not be a major concern. Extended quasi-likelihood is compared with Carroll & Ruppert's (1982) pseudo-likelihood method, which gives consistent estimates and, when slightly modified, asymptotically unbiased tests. We quantify the notion of a problem in which the amount of statistical information is large in each unit, showing in this instance that the two estimates are closely related and may be asymptotically equivalent in many important cases. However, in some cases outside the exponential family, an asymptotic bias can persist.

1. INTRODUCTION

Consider the following mean-variance model for observable data y:

$$E(y_i) = \mu_i = \mu_i(\beta) = f(x_i, \beta) \; ; \; var(y_i) = \sigma^2 g^2(\mu_i, z_i, \theta). \qquad (1.1)$$

Here, y_i is the ith response variable of N independent observations, (x_i, z_i) are associated vectors of covariates, f is the regression function, β is a p-vector of regression parameters, σ is a scale parameter, and g is the variance function with variance parameter θ (r × 1). For example, the variance may be modeled as proportional to an unknown power of the mean:

$$g(\mu_i, z_i, \theta) = \mu_i^\theta, \qquad \mu_i > 0. \qquad (1.2)$$

Special cases of (1.1) are used in applications such as radioimmunoassay, econometrics, and chemical kinetics. Model (1.1) includes the class of generalized linear models, see McCullagh & Nelder (1983).

A usual aim is the estimation of β, with estimation of the variance function parameters as an adjunct. However, as discussed by Davidian & Carroll (1987) and Davidian, Carroll & Smith (unpublished), estimation of the variance function, in particular the parameter θ, is an important problem both for estimation of β and in its own right.

Most methods for estimating θ are "regression" methods based on generalized least squares. In these techniques, θ and σ are estimated by a weighted regression of some function of the absolute residuals from a fit $\hat{\beta}_*$ on their expectations. For example, in location-scale problems squared residuals have approximate mean proportional to $g^2(\mu_i, z_i, \theta)$ and variance proportional to $g^4(\mu_i, z_i, \theta)$. Thus an estimate of θ can be obtained by a

generalized least squares regression of squared residuals on $\sigma^2 g^2(\hat{\mu}_i, z_i, \theta)$ with variance function $g^4(\hat{\mu}_i, z_i, \theta)$, where $\hat{\mu}_i = f(x_i, \hat{\beta}_*)$. A related method is the pseudo-likelihood approach of Carroll & Ruppert (1982). In this method, one pretends $\beta = \hat{\beta}_*$ and then estimates (σ, θ) by normal theory maximum likelihood, maximizing $\ell_{PL}(\hat{\beta}_*, \theta, \sigma)$, where

$$
\begin{aligned}
\ell_{PL}(\beta, \theta, \sigma) = {}& -N \log \sigma - \sum_{i=1}^{N} \log \left[g\{\mu_i(\beta), z_i, \theta\} \right] \\
& - (2\sigma^2)^{-1} \sum_{i=1}^{N} \{y_i - f(x_i, \beta)\}^2 / g^2\{\mu_i(\beta), z_i, \theta\} .
\end{aligned} \tag{1.3}
$$

This process may be iterated with a generalized least squares routine for β. The number of iterations of the entire procedure for estimation of β may be chosen in advance or the process may be iterated until convegence; see Davidian & Carroll (1987). The pseudo-likelihood method is asymptotically equivalent to weighted regression on squared residuals with estimated weights, and full iteration of such a regression yields the pseudo-likelihood estimate. Both methods can be modified to account for loss of degrees of freedom for preliminary estimation of β as in Harville (1977); for a discussion and a review of many common methods for estimation of θ, see Davidian & Carroll (1987).

Pseudo-likelihood and weighted squared residual estimation are based upon the method of moments. Nelder & Pregibon (1987) instead attempt to define a family of distributions with mean and variance functions given by (1.1), this class including as special cases skewed distributions such as the Poisson or gamma. Their extended quasi-likelihood is

$$\ell_{QL}(\beta,\theta,\sigma) = (-1/2) \sum_{i=1}^{N} [\log \{2\pi\sigma^2 g^2(y_i,z_i,\theta)\} + D\{y_i,\mu_i(\beta),z_i,\theta\}/\sigma^2],$$

where (1.4)

$$D(y,\mu,z,\theta) = -2 \int_{y}^{\mu} \frac{y-w}{g^2(w,z,\theta)} dw.$$

The function ℓ_{QL} is sometimes but not always an exact log-likelihood. Under (1.2), if $\theta = 0$, ℓ_{QL} is the normal log-likelihood; for $\theta = 1.5$, that of the inverse Gaussian. For $\theta = .5$, $\sigma = 1$, ℓ_{QL} differs from the Poisson log-likelihood by replacing $y_i!$ by its Stirling approximation; for $\theta = 1$, ℓ_{QL} differs from the gamma log-likelihood by a factor depending on σ. One motivation for (1.4) is the Edgeworth expansion of Barndorff-Nielsen and Cox (1979) or the related saddlepoint approximation of Daniels (1954), which yield an expansion for the density of the mean of m random variables from a one parameter exponential family as $m \rightarrow \infty$. The leading term of the expansion at $m = 1$ is the extended quasi-likelihood summand. See Efron (1986) for a related formulation. Note that the form of ℓ_{QL} may be unsatisfactory in situations for which $g(y,z,\theta) = 0$ for $y = 0$. In this case Nelder & Pregibon suggest replacing $g(y,z,\theta)$ by $g(y+c,z,\theta)$ for some c; we use this adjustment where applicable in our discussion.

An additional reason for considering approximate likelihoods for a mean-variance model is that linear exponential families with given mean-variance relationship do not always exist. For example, Bar-Lev & Enis (1986) have shown that if the distribution of y_i is an exponential family with variance function (1.2), it is necessary that $\theta \notin (-\infty,0) \cup (0,1/2)$, so that such a family exists only when $\theta \in \{0\} \cup [1/2,\infty)$, and the general form for the density parameterized in terms of θ and σ is unwieldy.

We have observed in many examples that the pseudo-likelihood and quasi-likelihood methods lead to similar estimates, although sometimes inferences for θ are substantially different. In Section 2, we construct an asymptotic theory for extended quasi-likelihood which allows an easy illustration of the relationship between the two methods and suggests a simple motivation for the form of extended quasi-likelihood. In general, the extended quasi-likelihood estimate of θ is inconsistent, and the resulting test is biased, but in our experience, the inconsistency has not been major at exponential families. The inconsistency is noted independently by Morton (1987), who, like Carroll & Ruppert (1982), uses an estimate of θ based on squared residuals. These methods have the advantage of being consistent and, when slightly modified, asymptotically unbiased for testing. We study extended quasi-likelihood in the case that it is likely to perform best, namely, (Morton, 1987) when the amount of statistical information is large in each observation. We quantify this notion, and then show that in this instance the two estimators are nearly asymptotically equivalent, although extended quasi-likelihood can be affected by an asymptotic bias while pseudo-likelihood is not when the underlying distribution is asymmetric and outside the exponential family. In Section 3 we discuss inference for θ based on the two approaches. From the theory of Section 2 we observe that while inference based on asymptotic theory for the two approaches yields similar results under many conditions, such a test based on extended quasi-likelihood can be adversely affected by possible asymptotic bias of the estimator. The difference in test behavior we have observed may be due to the effect of asymptotic bias.

2. SOME ASYMPTOTIC RESULTS

Neither pseudo-likelihood nor extended quasi-likelihood are exact likelihood approaches. Pseudo-likelihood is based on the method of moments, so that the estimating equations are unbiased, and hence consistency and asymptotic normality obtain under very general conditions even without the assumption of normality. Let $v(\mu_i, z_i, \theta) = \log g(\mu_i, z_i, \theta)$, $v_\theta(\mu_i, z_i, \theta)$ be its column vector of partial derivatives with respect to θ, $\omega_\theta(\mu_i, z_i, \theta) = v_\theta(\mu_i, z_i, \theta) - N^{-1}\Sigma \, v_\theta(\mu_j, z_j, \theta)$, and $\xi(\mu, z, \theta) = \lim N^{-1}\Sigma \, \omega_\theta(\mu_i, z_i, \theta) \, \omega_\theta(\mu_i, z_i, \theta)^t$. Let subscripts denote differentiation with respect to the argument, e.g., $g_\mu(\mu_i, z_i, \theta) = \partial g(\mu_i, z_i, \theta)/\partial \mu_i$. Define the errors $\epsilon_i = (y_i - \mu_i)/\{\sigma \, g(\mu_i, z_i, \theta)\}$, and assume the $\{\epsilon_i\}$ are independent with skewness ζ_i and kurtosis κ_i; $\kappa_i = 0$ for normality. Let $\gamma = (\eta, \theta^t)^t$ and use subscripts PL and QL to denote pseudo-likelihood and extended quasi-likelihood, respectively.

RESULT 1 (Davidian and Carroll, 1987). Suppose that $(\hat{\beta}_* - \beta)/\sigma = O_p(N^{-1/2})$ and $\hat{\gamma}_{PL} - \gamma = O_p(N^{-1/2})$. Then $\hat{\theta}_{PL}$ is asymptotically normally distributed with mean θ. If $\sigma \to 0$ simultaneously with $N \to \infty$, then

$$N^{1/2}(\hat{\theta}_{PL} - \theta) = \{2\xi(\mu, z, \theta)\}^{-1} \, N^{-1/2} \sum_{i=1}^{N} (\epsilon_i^2 - 1) \, \omega_\theta(\mu_i, z_i, \theta) + o_p(1).$$

(2.1)

If the $\{\epsilon_i\}$ are identically distributed with kurtosis κ, the covariance matrix of the asymptotic distribution of $\hat{\theta}_{PL}$ is given by

$$(2 + \kappa) \, \{4N \, \xi(\mu, z, \theta)\}^{-1}.$$

(2.2)

The assumption $\sigma \to 0$ is a useful simplification technically and is relevant in applications where σ is "small" relative to the means as in assay problems, see Davidian, Carroll & Smith (unpublished). In the gamma and lognormal distributions, σ is the coefficient of variation, which is often fairly small. Alternatively, think of y_i as the mean of m observations with mean μ_i and variance $g^2(\mu_i, z_i, \theta)$, equate σ and $m^{-1/2}$, and let $m \to \infty$.

The assumption $\sigma \to 0$ yields a motivation for (1.4). Since the goal of extended quasi-likelihood is to describe a class of distributions "nearly" containing exponential families, consider a density h such that

$$\log h(y, \alpha, \theta, \sigma) = \{y\alpha - b(\alpha)\}/\sigma^2 + c(y, \theta, \sigma) \qquad (2.3)$$

for some b, c, and $\alpha = \alpha(\mu, z, \theta)$. To satisfy (1.1) we require $\partial b(\alpha)/\partial \alpha = \mu$ and $\partial^2 b(\alpha)/\partial \alpha^2 = g^2(\mu, z, \theta)$, implying that $\mu = \{\partial b(\alpha)/\partial \mu\} \; \partial \mu/\partial \alpha$ and $g^2(\mu, z, \theta) = \partial \mu/\partial \alpha$. This yields, writing b now as a function of μ,

$$\alpha = \int_{-\infty}^{\mu} \{1/g^2(u, z, \theta)\} \; du; \qquad b(\mu, z) = \int_{-\infty}^{\mu} \{u/g^2(u, z, \theta)\} \; du.$$

Plugging into (2.3) gives after simplification

$$\log h(y, \alpha, \theta, \sigma) = -(2\sigma^2)^{-1} D(y, \mu, z, \theta) + d(y, \theta, \sigma) \qquad (2.4)$$

for some function d. For h to be a density we must choose d so that h integrates to one; by approximating the first term on the right side of (2.4) when σ is small we may approximate d. Since when σ is small we have $\sigma^{-2} D(y, \mu, z, \theta) \approx (y-u)^2/\{\sigma^2 g^2(y, z, \theta)\}$, $d \cong -(1/2) \log\{2\pi\sigma^2 g^2(y, z, \theta)\}$.

Inserting this in (2.4) yields the summand of (1.4).

The fact that (1.4) is an approximate log-likelihood implies that $\hat{\theta}_{QL}$ need not be consistent. With the suggested adjustment c = 1/6 as in Nelder & Pregibon (1987), if the $\{y_i\}$ are distributed as Poisson with means $\{\mu_i\}$ taking on values 1 and 4 in equal proportions and we assume $\theta < 1.0$, the theory of M-estimation as in Huber (1981 p. 130-132) implies that $\hat{\theta}_{QL}$ converges to 0.45. If the $\{\mu_i\}$ take on larger values, such as 30, 40 and 50 in equal proportions, however, $\hat{\theta}_{QL}$ converges to 0.50. Many examples in regression we have seen suggest that extended quasi-likelihood and pseudo-likelihood are typically equivalent for power of mean models (1.2).

Since the estimating equation for the extended quasi-likelihood estimate $\hat{\theta}_{QL}$ can be biased, standard asymptotic theory for $\hat{\theta}_{QL}$, while possible to construct, is not fully informative. As an approximation we use the small σ assumption to construct an asymptotic theory. We also describe an approach suggested by the Poisson case for "large" $\{\mu_i\}$.

RESULT 2. Suppose that $N^{1/2}(\hat{\gamma}_{QL} - \gamma) = 0_p(1)$ and $N^{1/2}(\hat{\beta}_{QL} - \beta)/\sigma = 0_p(1)$ if $N^{1/2}\sigma \to \lambda \geq 0$ as $N \to \infty$, $\sigma \to 0$. Then

$$N^{1/2}(\hat{\theta}_{QL} - \theta) = \{2 \, \xi(\mu,z,\theta)\}^{-1} N^{-1/2} \sum_{i=1}^{N} (\epsilon_i^2 - 1) \, \omega_\theta(\mu_i,z_i,\theta)$$

$$+ (N^{1/2}\sigma) \, \{6 \, \xi(\mu,z,\theta)\}^{-1} C_N + o_p(1), \text{ where} \qquad (2.5)$$

$$C_N = N^{-1} \sum_{i=1}^{N} \zeta_i \{g(\mu_i,z_i,\theta) \, v_{\theta\mu}(\mu_i,z_i,\theta) - 2 \, g_\mu(\mu_i,z_i,\theta) \, \omega_\theta(\mu_i,z_i,\theta)\}.$$

A sketch of the proof is contained in the Appendix. The implication of (2.5) is that while $\hat{\theta}_{PL}$ and $\hat{\theta}_{QL}$ behave similarly, they differ in an asymptotic fashion through the second term on the right hand side of (2.5)

in a way that might affect asymptotic inference. When $\hat{\theta}_{QL}$ is asymptotically normal it will have mean $\theta + \{6N^{-1/2}\xi(\mu,z,\theta)\}^{-1}\{\lambda C\}$, where $C_N \to C$. From (2.2) and (2.5), then, $\hat{\theta}_{QL}$ and $\hat{\theta}_{PL}$ will be asymptotically equivalent only if $\lambda = 0$ or $C_N \to 0$; the latter will occur for symmetrically distributed data. If the $\{\epsilon_i\}$ are identically distributed with kurtosis κ, for example, then $\hat{\theta}_{QL}$ and $\hat{\theta}_{PL}$ will have asymptotic covariance (2.2).

In the case of (1.2), $v_\theta(\mu_i, z_i, \theta) = \log \mu_i$ so that

$$C_N = N^{-1} \sum_{i=1}^{N} \zeta_i \mu_i^{\theta-1} \{1 - 2\theta (\log \mu_i - \overline{I}_N)\}, \quad \overline{I}_N = \sum_{j=1}^{N} \log \mu_j.$$

For the normal distribution, $\zeta_i = \kappa_i = 0$; for the gamma, lognormal, and inverse Gaussian distributions $\zeta_i = O(\sigma)$ and $\kappa_i = O(\sigma^2)$, so that the asymptotic bias is 0 and the two estimators are asymptotically equivalent with covariance the same as if the data were normally distributed with mean μ_i and variance $\sigma^2 \mu_i^{2\theta}$. From Bar-Lev & Enis (1986), $\zeta_i = O(\sigma)$ for distributions which are exponential families with $\theta \in \{0\} \cup [1/2, \infty)$. If the $\{y_i\}$ are not from an exponential family, the asymptotic bias need not be zero. For example, consider a shifted gamma model $y_i = \mu_i + \sigma g(\mu_i, z_i, \theta)\epsilon_i$, where w_i has a gamma (α_i, φ_i) distribution with $E(w_i) = \alpha_i/\varphi_i$, and $\epsilon_i = \{w_i - (\alpha_i/\varphi_i)\}(\alpha_i/\varphi_i^2)^{-1/2}$, so that $E(\epsilon_i^2) = 1$. In this case $\zeta_i = 2\alpha_i^{-1/2}$, so that if the $\{\alpha_i\}$ do not depend on σ, the asymptotic bias will not vanish. At exponential family models in cases not covered by the asymptotic theory here, one might expect pseudo-likelihood to be more variable than extended quasi-likelihood, since the latter is based on approximate exponential family likelihoods.

An asymptotic theory for which the means are "large" in which σ remains fixed yields a similar result under (1.2) if $\theta < 1$. Let $\mu_{0,N}$ be a sequence to be chosen shortly. Define $\mu_i^* = \mu_i/\mu_{0,N}$ and $y_i^* = y_i/\mu_{0,N}$ so that $\epsilon_i = \{y_i^* - \mu_i^*\}/(\delta \mu_i^{*\theta})$, where $\delta = \sigma \mu_{0,N}^{(\theta-1)}$. If as $N \to \infty$, $\min_i \mu_i \to \infty$ and $\mu_{0,N} \to \infty$ in such a way that the $\{\mu_i^*\}$ and the $\{y_i^*\}$ are well-behaved, then if $\theta < 1$, $\delta \to 0$ as $N \to \infty$ so that the calculations here parallel those for the case of small σ. By analogy, the small σ part of Result 1 holds. Replacing σ by δ in (2.5), in the Poisson case for which $\theta = .5$ and $\sigma = 1$, $\zeta_i = \mu_i^{-1/2}$ and $\kappa_i = \mu_i^{-1}$, so that $C_N \to 0$ and the limiting covariance of $\hat{\theta}_{QL}$ is as if $\kappa_i = 0$. Thus, in the case of "large" means and data distributed as Poisson, extended quasi-likelihood and pseudo-likelihood will behave similarly.

The theory presented here is applicable when the small σ or large mean assumption is valid, which is the case in many important situations, and does not address problems of other types.

3. INFERENCE FOR θ

The asymptotic distribution theory of Section 2 can be used to construct tests of H_0: $\theta = \theta_0$. Throughout, define $\Lambda(\mu,z,\theta,\kappa) = \lim N^{-1}\Sigma$ $(2+\kappa_i)\ \omega_\theta(\mu_i,z_i,\theta)\ \omega_\theta(\mu_i,z_i,\theta)^t$. From (1.3), $\hat{\theta}_{PL}$ maximizes $\ell_{PL}^*(\hat{\beta}_*,\theta)$, where

$$\ell_{PL}^*(\beta,\theta) = -N \log \hat{\sigma}_{PL}(\beta,\theta) - \sum_{i=1}^{N} \log g\{\mu_i(\beta),z_i,\theta\},$$

$$\hat{\sigma}_{PL}^2(\beta,\theta) = \sum_{i=1}^{N} \{(y_i - f(x_i,\beta))\}^2/g^2\{\mu_i(\beta),z_i,\theta\}.$$

One might reasonably base inference for θ on a test statistic

$$T_N = -2 \, [\ell_{PL}^{*}\{\hat{\beta}(\theta_0),\theta_0\} - \ell_{PL}^{*}\{\hat{\beta}(\hat{\theta}_{PL}),\hat{\theta}_{PL}\}],$$

where $\hat{\beta}(\theta)$ denotes a generalized least squares estimate computed at θ and compare T_N to the percentiles of the χ_r^2 distribution. Although $\hat{\beta}(\hat{\theta}_{PL})$ and $\hat{\theta}_{PL}$ do not necessarily jointly maximize the pseudo-likelihood, the fact that $\{\hat{\beta}(\hat{\theta}_{PL}) - \hat{\beta}_{*}\}$ converges in probability to 0 along with a Taylor series and Result 1 may be used to show that under H_0, T_N has asymptotically the same distribution as the random variable $2 \, W(0)^{t}\xi(\mu,z,\theta)W(0)$, where $W(M)$ has a normally distribution with mean M and covariance matirix $\{\xi(\mu,z,\theta)\}^{-1}\Lambda(\mu,z,\theta)\{\xi(\mu,z,\theta)\}^{-1}/4$.

Nelder & Pregibon (1987) suggest a likelihood ratio type test based on treating the extended quasi-likelihood as an actual likelihood. Such a test is based on

$$Q_N = -2 \, [\ell_{QL}^{*}\{\hat{\beta}(\theta_0),\theta_0\} - \ell_{QL}^{*}\{\hat{\beta}(\hat{\theta}_{QL}),\hat{\theta}_{QL}\}],$$

where

$$\ell_{QL}^{*}(\beta,\theta) = -N \, \log \, \sigma_{QL}(\beta,\theta) - \sum_{i=1}^{N} \, \log \, g(y_i,z_i,\theta),$$

$$\sigma_{QL}^2(\beta,\theta) = N^{-1} \sum_{i=1}^{N} \, D\{y_i,\mu_i(\beta),z_i\}.$$

In the situation of Result 2, Q_N has asymptotically the same distribution as the random variable $2W[\lambda C\{6\xi(\mu,z,\theta)\}^{-1}]^{t}\Lambda(\mu,z,\theta)W[\lambda C\{6\xi(\mu,z,\theta)\}^{-1}]$ under H_0, where $C_N \rightarrow C$.

The asymptotic distributions of T_N and Q_N need not be χ_r^2 in general. Appropriately scaled versions of these statistics, say aT_N and aQ_N for some constant a, will be possibly noncentral χ_r^2 if and only if $\Lambda(\mu,z,\theta) = $

$2\xi(\mu,z,\theta)/a$, see, for example, Muirhead (1982, p. 31, Theorem 1.4.5). Thus, comparison to the percentiles of the χ_r^2 distribution may be misleading. In many important special cases of practical application, however, the distributions are readily seen to be chi-square. If the distributions of the $\{\epsilon_i\}$ are normal, so that $\kappa_i \equiv 0$, then T_N and Q_N are both asymptotically χ_r^2. If the $\{\epsilon_i\}$ are identically distributed with kurtosis κ, since then $\Lambda(\mu,z,\theta) = (2 + \kappa) \xi(\mu,z,\theta)$, it follows that under H_0, $\{2/(2+\kappa)\} T_N$ is asymptotically distributed as χ_r^2, so that a test based on this statistic with κ appropriately estimated is an asymptotic α-level test. McCullagh & Pregibon (1987) consider estimators for the cumulants for linear regression models. In the situation of Result 2, under H_0, $\{2/(2+\kappa)\}$ is asymptotically distributed as noncentral χ_r^2 with noncentrality parameter $\Delta = \lambda^2 C^t \xi(\mu,\theta)^{-1} C \{9(2+\kappa)\}^{-1}$. As long as $\Delta = 0$, comparing this statistic to the percentiles of the χ_r^2 distribution is an asymptotic α-level test which is asymptotically equivalent to the test based on T_N.

For general, not necessarily identically distributed $\{\epsilon_i\}$, the asymptotic distribution may not necessarily be chi-square. However, if $r = 1$ so that θ, $\xi(\mu,z,\theta)$, and $\Lambda(\mu,z,\theta)$ are scalar, as in the important cases of extra variation in Poisson or binomial models, under H_0, $2\xi(\mu,z,\theta)/\Lambda(\mu,z,\theta)T_N$ is asymptotically distributed as χ_1^2, while $2\xi(\mu,z,\theta)/\Lambda(\mu,z,\theta)Q_N$ is asymptotically distributed as noncentral χ_1^2 with noncentrality parameter $\Delta = \lambda^2 C^2/\{9\Lambda(\mu,z,\theta)\}$. In practice, one might estimate this factor by computing appropriate estimates for $\xi(\mu,z,\theta)$ and $\Lambda(\mu,z,\theta)$. For example, if nearly Poisson data are suspected, one might estimate κ_i by the final estimate for μ_i^{-1}.

Nelder & Pregibon suggest comparing Q_N directly to the percentiles of the χ_r^2 distribution. Comparing either T_N or Q_N to the χ_r^2 distribution may

be erroneous in general, but in important practical situations such as those above in which the distributions are χ_r^2, their suggestion does not account for the additional multiplicative factor depending on kurtosis which appears unless the data are normal. In the saddlepoint approximation approach, $m \to \infty$ implies $\kappa_1 \approx 0$, thus they observe that if the underlying distribution of the data is known to be from an exponential family, then such a test is asymptotically valid. In our asymptotics, for the cases of the normal, gamma, and inverse Gaussian examples cited in Section 2 we see this to be the case. For the Poisson case, one may consider the analogous "large mean" asymptotic approach at the end of Section 2 to conclude the same. We further obtain the correct form and properties for a test of this type when only the mean-variance relationship is specified. For other approaches to variance function estimation which avoid problems of kurtosis, see Davidian & Carroll (1987) and Giltinan, Carroll, & Ruppert (1986).

For a model such as (1.1) for which only the mean and variance are specified, interest in θ may be in the context of trying to understand the structure of the variances, not the form of the underlying distributions. When appropriate, a chi-square test based on Q_N will approach its nominal level if $\Delta = 0$. When the underlying distributions of the data are such that $\hat{\theta}_{QL}$ is biased asymptotically so that $\Delta \neq 0$, the validity of a χ^2 test based on Q_N may be seriously affected.

ACKNOWLEDGEMENT

APPENDIX : SKETCH OF PROOF OF RESULT 2

For convenience, let $\eta = \log \sigma$ and let $(\hat{\beta}, \hat{\theta}, \hat{\eta})$ be the joint extended quasi-likelihood estimators for (β, θ, η). Let $\tau(1, \beta, \theta) = \{1, v_\theta^t(\mu_i, z_i, \theta)\}^t$. $\rho(1, \beta, \theta) = f_\beta(x_i, \beta)/g\{\mu_i(\beta), z_i, \theta\}$, $B_N = -4N^{-1}\Sigma \, \tau(1, \beta, \theta) \, \tau(1, \beta, \theta)^t$, and $G_N = N^{-1}\Sigma \, \rho(1, \beta, \theta) \, \rho(1, \beta, \theta)^t$. From (1.4), $(\hat{\beta}, \hat{\theta}, \hat{\eta})$ solves

$$0 = \begin{bmatrix} Q_{1,N}(\hat{\beta}, \hat{\theta}, \hat{\eta}) \\ Q_{2,N}(\hat{\beta}, \hat{\theta}, \hat{\eta}) \\ Q_{3,N}(\hat{\beta}, \hat{\theta}, \hat{\eta}) \end{bmatrix}, \quad \text{where} \tag{A.1}$$

$$Q_{1,N}(\beta, \theta, \eta) = N^{-1/2} \sum_{i=1}^{N} e^{-2\eta} (Y_i - \mu_i) \, \rho(1, \beta, \theta)/g(\mu_i, z_i, \theta);$$

$$Q_{2,N}(\beta, \theta, \eta) = N^{-1/2} \sum_{i=1}^{N} \{ e^{-2\eta} D(Y_i, \mu_i, z_i, \theta) - 1 \};$$

$$Q_{3,N}(\beta, \theta, \eta)$$
$$= N^{-1/2} \sum_{i=1}^{N} [-\tfrac{1}{2} e^{-2\eta} \, \partial D(Y_i, \mu_i, z_i, \theta)/\partial\theta - \{ \partial \log g(Y_i, z_i, \theta)/\partial\theta \}].$$

The following result is shown by assuming appropriate smoothness conditions for g so that D may be differentiated.

LEMMA A. Under regularity conditions,

$$N^{-1} \sum_{i=1}^{N} D(Y_i, \mu_i, z_i, \theta) = \sigma^2 \, N^{-1} \sum_{i=1}^{N} \epsilon_i^2 + \sigma^3 \, N^{-1} \sum_{i=1}^{N} s_{1,i} \, \epsilon_i^3 + 0_p(\sigma^4);$$

$$N^{-1} \sum_{i=1}^{N} \partial D(Y_i, \mu_i, z_i, \theta)/\partial\theta = -2 \{ \sigma^2 N^{-1} \sum_{i=1}^{N} \epsilon_i^2 v_\theta(i, \beta, \theta)$$

$$+ \sigma^3 N^{-1} \sum_{i=1}^{N} s_{2,i} \epsilon_i^3 \} + 0_p(\sigma^4);$$

$$N^{-1} \sum_{i=1}^{N} \partial^2 D(Y_i, \mu_i, z_i, \theta)/\partial\theta\partial\theta^t = 2 \sigma^2 N^{-1} \sum_{i=1}^{N} \epsilon_i^2 [3 v_\theta(i, \beta, \theta) v_\theta^t(i, \beta, \theta)$$

$$-\{g_{\theta\theta}(\mu_i, z_i, \theta) / g(\mu_i, z_i, \theta)\}] + 0_p(\sigma^3).$$

where $g_{\theta\theta}(\mu_i, z_i, \theta) = \partial^2\{g(\mu_i, z_i, \theta)\}/\partial\theta\partial\theta^t$, $s_{1,i} = -2 g_\mu(\mu_i, z_i, \theta)/3$, and $s_{2,1} = g_{\theta\mu}(\mu_i, z_i, \theta)/3 - v_\theta(\mu_i, z_i, \theta) g_\mu(\mu_i, z_i, \theta)$.

A Taylor series in (A.1) using consistency, a Taylor series in σ about 0 using Lemma A and laws of large numbers yield after simplification

$$\begin{bmatrix} G_N & 0 \\ 0 & \frac{1}{2} B_N \end{bmatrix} N^{1/2} \begin{bmatrix} (\hat{\beta} - \beta)/\sigma \\ \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix}$$

$$\text{(A.2)}$$

$$= \begin{bmatrix} N^{-1/2} \sum_{i=1}^{N} \epsilon_i \, \rho(i, \beta, \theta) \\ N^{-1/2} \sum_{i=1}^{N} (\epsilon_i^2 - 1) \, \tau(i, \beta, \theta) \end{bmatrix} + (N^{1/2}\sigma) \begin{bmatrix} 0 \\ N^{-1} \sum_{i=1}^{N} \epsilon_i^3 s_i \end{bmatrix} + o_p(1),$$

where $s_i^t = (s_{1,i}, s_{2,i}^t)$. Equation (A.2) implies that, as $N \to \infty$, $\sigma \to 0$,

$$B_N N^{1/2} \begin{bmatrix} \hat{\eta} - \eta \\ \hat{\theta} - \theta \end{bmatrix} = 2 N^{-1/2} \sum_{i=1}^{N} (\epsilon_i^2 - 1) \tau(i, \beta, \theta)$$

$$+ 2 (N^{1/2} \sigma) N^{-1} \sum_{i=1}^{N} \epsilon_i^3 s_i + o_p(1).$$

Algebra and simple probability limit calculations yield the result. Equation (A.2) also shows that in these asymptotics $\hat{\beta}_{QL}$ is equivalent to a generalized least squares estimator for β. $\square$

REFERENCES

Bar-Lev, S.K. & Enis, P. (1986). Reproducibility and natural exponential families with power variance functions. Annals of Statistics 14, 1507-1522.

Barndorff-Nielsen, O. and Cox, D.R. (1979). Edgeworth and saddle-point approximations with statistical applications. Journal of the Royal Statistical Society, Series B 41, 279-312.

Carroll, R.J. & Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. Annals of Statistics 10, 429-441.

Daniels, H.E. (1954). Saddlepoint approximations in statistics. Annals of Mathematical Statistics 25, 631-650.

Davidian, M. & Carroll, R.J. (1987). Variance function estimation. *Journal of the American Statistical Association*, to appear.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* 81, 709-721.

Giltinan, D.M., Carroll, R.J., & Ruppert, D. (1986). Some new methods for weighted regression when there are possible outliers. *Technometrics* 28, 219-230.

Harville, D. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 79, 302-308.

Huber, P.J. (1981) *Robust Statistics*. New York: John Wiley and Sons.

McCullagh, P. & Nelder, J.A. (1983). *Generalized Linear Models*. New York: Chapman and Hall.

McCullagh, P. & Pregibon, D. (1987). k-statistics in regression. *Annals of Statistics*, to appear.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: John Wiley and Sons.

Nelder, J.A. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* 74, 221-232.